

[오픈소스 프로젝트 소개]

아태지역 최초 NVIDIA DGX-Ready Software 인증 'Backend.AI'

('26.3.20, 작성 : 래블업 김준기 CTO)

Backend.AI는 AI 개발 및 고성능 컴퓨팅을 위한 GPU/AI 가속기 벤더 독립적 워크로드 호스팅 플랫폼으로, 자체 오케스트레이션 엔진과 작업 스케줄러를 기반으로 클라우드와 온프레미스 환경에서 AI 워크로드를 효율적으로 운영할 수 있도록 지원하는 오픈소스 프로젝트이다.

사용자와 컨트리뷰터는 GitHub 주 프로젝트 저장소를 중심으로 Backend.AI Everywhere Discord 커뮤니티를 통해 설치, 운영, 트러블슈팅, 활용 방법 등을 공유하며 활발한 기술 교류를 진행하고 있다.

이 프로젝트는 래블업(Lablup)에서 기술지원을 하고 있으며, 래블업의 공동창립자이자 CTO인 김준기 개발자가 참여하고 있으며, AI 인프라와 고성능 컴퓨팅 분야에서 다양한 오픈소스 활동을 이어가고 있다.


□ 오픈소스 허브 프로젝트 - Backend.AI

▶ 오픈소스 프로젝트(OpenSource Project) - Backend.AI

구분	설명
프로젝트 개요	프로젝트 저장소 https://github.com/lablup/backend.ai
	홈페이지 https://www.lablup.com/
	라이선스 LGPLv3 (코어 엔진) / MIT License (클라이언트 및 공통 라이브러리)
	프로젝트 분야 Cloud
	프로젝트 소개 AI 개발 및 고성능 컴퓨팅을 위한 GPU/AI 가속기 벤더 독립적 워크로드 호스팅 플랫폼으로, 자체 개발한 오케스트레이션 엔진과 작업 스케줄러를 기반으로 클라우드 및 온프레미스/폐쇄망 환경에서의 클러스터 운영 기술을 백엔드부터 프론트엔드까지 모두 경험할 수 있는 프로젝트
핵심 가치	<ul style="list-style-type: none"> - 컨테이너 오케스트레이션을 위한 개별 컨테이너의 수명주기 관리부터 클러스터 수준의 스케줄링 추상화 계층까지 전체 스택을 모두 보유하여 다중노드 분산시스템 관점에서 수직통합을 경험해볼 수 있는 프로젝트 - Slurm과 Kubernetes의 장점을 조합하고 빠진 부분을 채워넣는 방식으로 자체 개발하였기 때문에 오픈소스 생태계에서 고유한 클러스터 수준 작업 스케줄러 대안 구현체로서 의의가 있음 * (편집자 주) Slurm(Simple Linux Utility for Resource Management)은 리눅스 클러스터 환경에서 CPU, GPU, 메모리 등의 자원을 효율적으로 관리하고 작업 스케줄링을 수행하는 오픈소스 워크로드 매니저 * (편집자 주) Kubernetes(쿠버네티스)는 컨테이너화된 애플리케이션을 자동으로 배포, 확장(scaling), 관리하는 오픈소스 시스템

구분	설명	
주요 특징	등장배경	<ul style="list-style-type: none"> - 클라우드 환경에서 소프트웨어 배포 편의성과 재현 가능성을 개선하는 컨테이너 기술의 부상 - 일반적으로 AI/HPC 워크로드는 호스트의 모든 연산 자원을 점유한다는 가정 하에 작성되는데, 컨테이너 환경에서 자원 제약이 설정되어 있더라도 하드웨어의 부분적 가시성 제약이 보장되지 않아 연산자원 과다 경합 발생 - Slurm은 매우 고도화된 스케줄링 옵션을 지원하지만 단일 노드를 여러 워크로드가 공유하기 어려움 - Kubernetes는 CPU-메모리만 사용하면서 언제든지 재구동할 수 있는 워크로드를 잘 다루지만 stateful 하면서 하드웨어 구성(토폴로지 등)에 민감한 워크로드를 다루기 어려움 <p>* (편집자 주) stateful : 상태정보 유지</p>
	핵심 기능	<ul style="list-style-type: none"> - 다종·이종 AI 가속기에 대한 연산자원 추상화 계층 - 토폴로지 및 워크로드 밀도 기반 작업 배치 및 노드 내 리소스 매핑 - 장치 단위 또는 분할 가상화된 가속기 자원을 할당·관리 할 수 있는 스케줄러 계층 - 배치·추론·인터랙티브 워크로드 유형을 단일 자원 풀에서 혼합 실행하거나 논리적 자원풀을 통한 접근제어 및 정책 기반 관리 - Prometheus, OpenTelemetry 기반 상세 모니터링 통합 <p>* (편집자 주) Prometheus, OpenTelemetry : 오픈소스 모니터링 및 분석 솔루션</p> <ul style="list-style-type: none"> - 스토리지 추상화 계층을 통한 제조사 특화 파일시스템 가속 API 지원 (예: GPUDirect Storage)
	차별화 요소	<ul style="list-style-type: none"> - 데스크톱 규모부터 데이터센터 규모까지 단일 아키텍처로 스케일링 - 아태지역 최초·유일 NVIDIA DGX-Ready Software 인증 - GS 1등급 인증
	대상 사용자	<ul style="list-style-type: none"> - 여러 연구원이 공유해야 하는 GPU 인프라를 정책 기반으로 관리 자동화를 하고 싶은 사용자 - 학습과 추론 워크로드를 동시에 또는 단독으로 실행 하면서 손쉽게 자원 배분을 동적으로 변경하고 성능 모니터링을 하고 싶은 사용자
	운영 환경	<ul style="list-style-type: none"> - Linux (Ubuntu 22.04+, RHEL/CentOS 7+, Alma/Rocky 8+)/ macOS 11+ - Intel x86-64 / ARM64 (8+) - NVIDIA GPU (CUDA 10+ 지원), AMD GPU (ROCm 6+ 지원), 기타 컨테이너환경 지원 가능한 AI 가속기
	활용 분야	<ul style="list-style-type: none"> - AI 워크로드(훈련·미세조정·추론·개발) 실행·관리 - HPC 워크로드(시뮬레이션·수치해석·통계) 실행·관리
	프로젝트 생태계	커뮤니티 현황 <ul style="list-style-type: none"> - GitHub (Backend.AI 주 저장소) - Discord (Backend.AI Everywhere)
오픈소스 허브	https://www.oss.kr/opensource/hub/56925	

▶ 커뮤니티(Community) - Backend.AI Everywhere

구분	설명	
커뮤니티 개요	커뮤니티 소개	<ul style="list-style-type: none"> - Backend.AI 개발자·기여자 간 교류 및 오픈소스 기여자 지원 - Backend.AI에 기여하고 있는 컨트리뷰터들과 사용자들이 설치 및 운영 관련 질문, 트러블슈팅, 활용 팁 등 다양한 정보를 공유하는 소통 채널입니다.
	홈페이지 및 SNS	https://discord.gg/zv3Pbtmpxv
	프로젝트 저장소	https://github.com/lablup/backend.ai
	커뮤니티 분야	Backend.AI 코어 개발, 플러그인 및 확장 제안, AI·MLOps 일반 관심 주제
	전문분야	Python asyncio, Rust 활용한 분산시스템 프로그래밍 / React, GraphQL 기반 프론트엔드 엔지니어링 / AI 및 MLOps 기술 전반
	커뮤니티 연혁	2024년 컨트리뷰션 아카데미 Raftify 프로젝트 호스팅 계기로 시작
	커뮤니티 미션/비전	Backend.AI 사용자 및 개발자 간 교류 확산
	글로벌 커뮤니티	현재 단독 커뮤니티로 운영 중
커뮤니티 활동	주요 활동	Backend.AI 사용자 교류, Raftify 등 컨트리뷰션 아카데미 활동 지원
	최근 활동 사례	 <p data-bbox="810 1783 1321 1816">그림 1. 2024년도 Raftify 멘토링 세션</p>
	커뮤니티 문화	자유 참여
커뮤니티 참여 방법	GitHub 이슈 발행 또는 Discussion 보드를 통한 질의응답, Discord 커뮤니티 참여	

▶ 주요 개발자 (Main Developers) - 김준기

구분	내용	
개발자 소개	이메일 또는 SNS	- joongi@lablup.com - https://www.facebook.com/achimnol
	전문 분야	aiotools, aio-libs/aiomonitor 메인테이너
	경력	現) 래블업 주식회사 공동창립자, CTO 前) Microsoft Research Cambridge, Research Intern KAIST 전산학 박사 2017-2019 글로벌 오픈프론티어 2024 ICT기술 사업화 페스티벌 과학 기술정보통신부 장관상
참여 프로젝트	https://github.com/lablup/backend.ai https://github.com/aio-libs/aiomonitor https://github.com/aio-libs/aiodocker https://github.com/achimnol/aiotools	
기여 내용	초기 MVP 단계에 이를 때까지 대부분의 설계·구현 수행, 현재는 프로젝트 운영에 집중	
오픈소스 참여 계기	재현가능한 계산 중심의 연구개발 환경을 만들고자 하는 문제를 해결하기 위해 컨테이너와 클라우드 기술을 결합하여 직접 프로젝트를 시작함	
기여 활동	대외 홍보, 전체 아키텍처 설계 및 로드맵 수립	
성장 포인트	처음에는 거의 1인 프로젝트 수준으로 백지에서 시작하였지만, 나중에는 10명 이상의 상시 개발자 및 전문 기술지원 팀을 꾸릴 정도로 상용화·제품화에 성공하며 개발과 프로젝트 운영 전 주기를 경험해볼 수 있었음	
후배들에게 조언	오픈소스 이력이나 경력을 만들어야 한다는 생각으로 접근하기 보다는, 실제로 본인이 맞닥뜨리고 해결하고 싶은 부분에 대해 명확하게 정의 가능한 문제를 중심으로 관련된 오픈소스를 찾아서 기여하거나 직접 오픈소스를 만들어보는 식으로 시작하는 것이 좋습니다. 이제 AI의 발전으로 다양한 기술적 아이디어들을 빠르게 프로토타이핑하고 실험해보기 좋아졌지만, 오픈소스는 여전히 기여자들끼리의 신뢰를 기반으로 움직입니다. 따라서 어떤 방법과 도구를 사용하든 항상 최고 품질의 코드를 지향하면서 적극적인 커뮤니케이션을 유지하는 것이 좋겠습니다.	